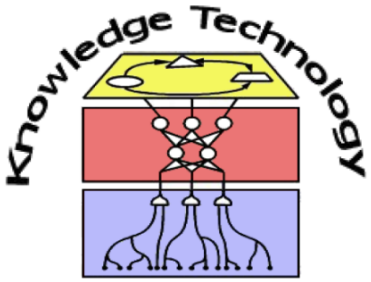
Comparison of Neural Architectures for Sentiment Analysis of Movie Reviews

Sören Hübner and Nicolás Pérez de Olaguer



http://www.informatik.uni-hamburg.de/WTM/

Motivation

- People search for and are affected by Online opinions
  - i.e. Trip Advisor, Goodreads, Amazon, Yelp,…
- Internet Movie Database (IMDb)
- Over 5 million titles
- Over 80 million users

- Would not be useful to see if a movie review is positive or negative?
- How? **Sentiment Analysis** using **Natural Language Processing** (NLP)

Outline

- **1**. Motivation and Question
- **2**. Basics and Definition
- **3**. Approach
- **4**. Results
- **5**. Discussion
- **6**. Conclusion

2. Sentiment Analysis

- An easy approach would be:
  - Count number of positive and negative words.
  - If: Positive > Negative = Positive Review
  - Else: Negative Review
  - Other option would be weighting keywords.
    - i.e. amazing = +2, awful = -3
  - Not so easy… Imagine this example:
    - *I used to hate the topic but this stuff is yuummy!* ☺
  - Let us find something more elaborated!

2. Sentiment Analysis

- Current state-of-the-art is word model creation.
  - Finding temporal relations, rhetorical uses, …
  - Can deal with: **Polarity flippers**, **multiword expressions** (slang), **subtle sentiments, …**
- Different examples like:
  - **Word2vec**
  - **Bag of Words**
  - **Embedding vector**

3. Approach

- Import IMDB review corpus. Properly labelled.
- Data pre-processing.
- Create a word model (embedding vector).
- Train different Neural Networks approaches for binary classifications.
  - Multi Layer Perceptron (MLP)
  - Convolutional Neural Network (CNN)
  - Long Short-Term Memory (LSTM)

3. Approach - Dataset

- Imported from **Keras** library. In this case, already pre-processed.
- Dataset of 25,000 movies reviews.
- Each review is encoded as a **sequence** of word indexes (integers).
- Indexed by overall **frequency** in the dataset.
- Division between **Training** and **Test** Set.

3. Approach – Dataset example

Review	Sentiment
”I dont know why people think this is such a bad movie. Its got a pretty good plot, some good action, and the change of location for Harry does not hurt either. Sure some of its offensive and gratuitous but this is not the only movie like that. Eastwood is in good form as Dirty Harry, and I liked Pat Hingle in this movie as the small town cop. If you liked DIRTY HARRY, then you should see this one, its a lot better than THE DEAD POOL. 4/5”	Positive
”I watched this video at a friend’s house. I’m glad I did not waste money buying this one. The video cover has a scene from the 1975 movie Capricorn One. The movie starts out with several clips of rocket blow-ups, most not related to manned flight. Sibrel’s smoking gun is a short video clip of the astronauts preparing a video broadcast. He edits in his own voice-over instead of letting us listen to what the crew had to say. The video curiously ends with a showing of the Zapruder film. His claims about radiation, shielding, star photography, and others lead me to believe is he extremely ignorant or has some sort of ax to grind against NASA, the astronauts, or American in general. His science is bad, and so is this video.”	Negative

- Clear examples of positive and negative reviews.

3. Approach – Embedding Vector

- We first used Word2vec model. Gave us decent to good results.

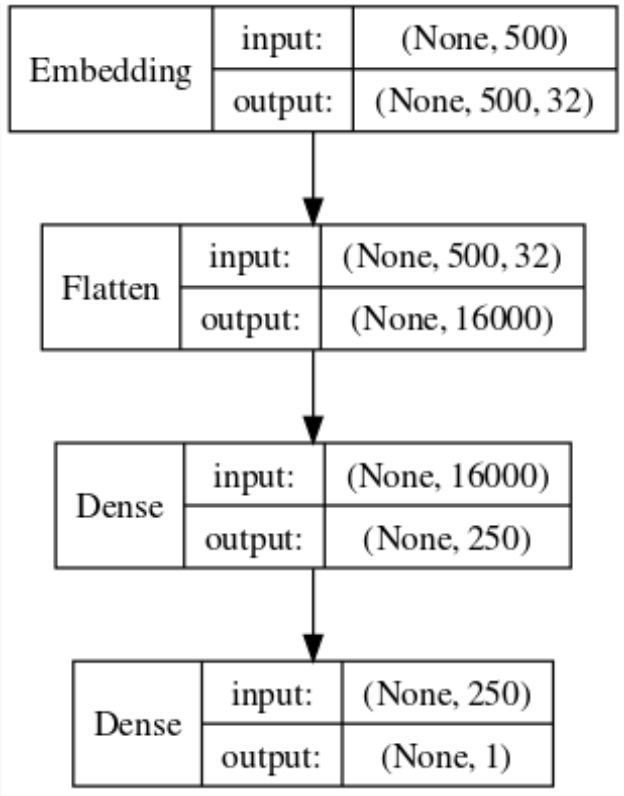
- We tried Keras embedding vector. Similar concept but not the same. [1, 2]

- Gave us better and faster results.

- We are ready to train!

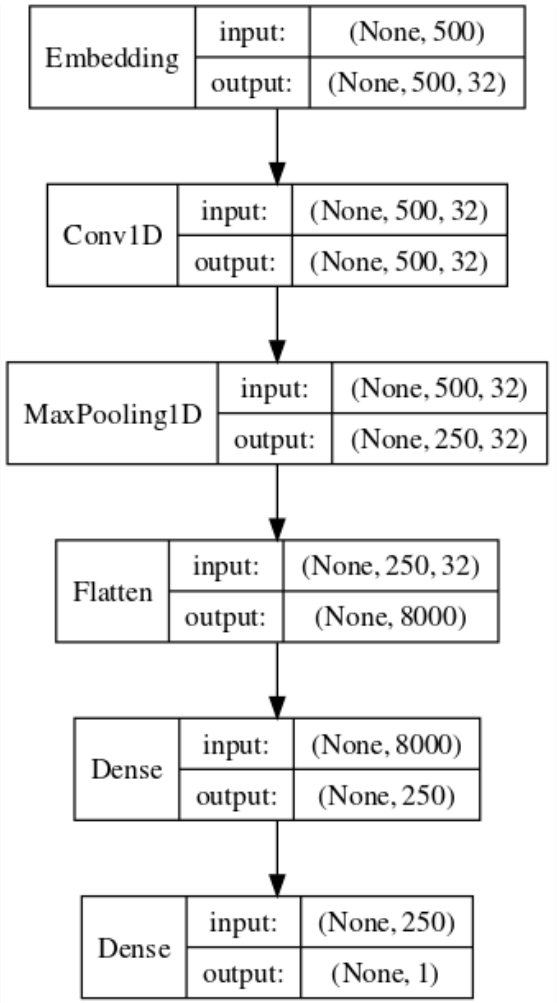
3. Approach – MLP

- "Multi-Layered Perceptron"
- Simpler ANN approach. Useful for binary classification.
- 250 fully connected layer + ReLu activation function.
- As good as other more complex architectures with simple tasks.



3. Approach – CNN

- Convolutional Neural Network useful for feature learning.
- Convolution always followed by a max pooling layer.
- In the end of the layer stack, we have a MLP for binary classification.
- Why? Learn a higher level of abstraction of the data.



3. Approach - LSTM

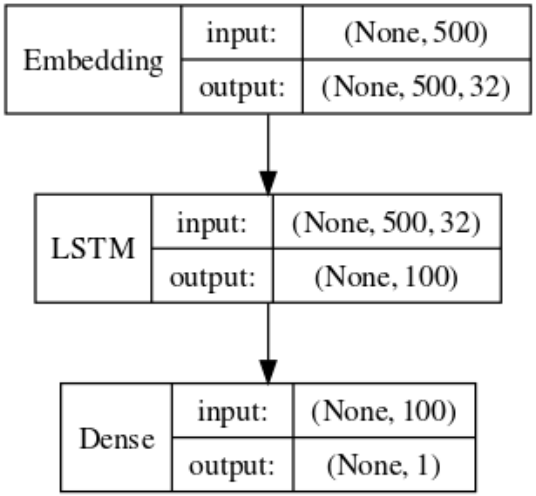
- "Long Short-Term Memory" type of Recurrent Neural Network (RNN).

- Captures time dependencies of the data.

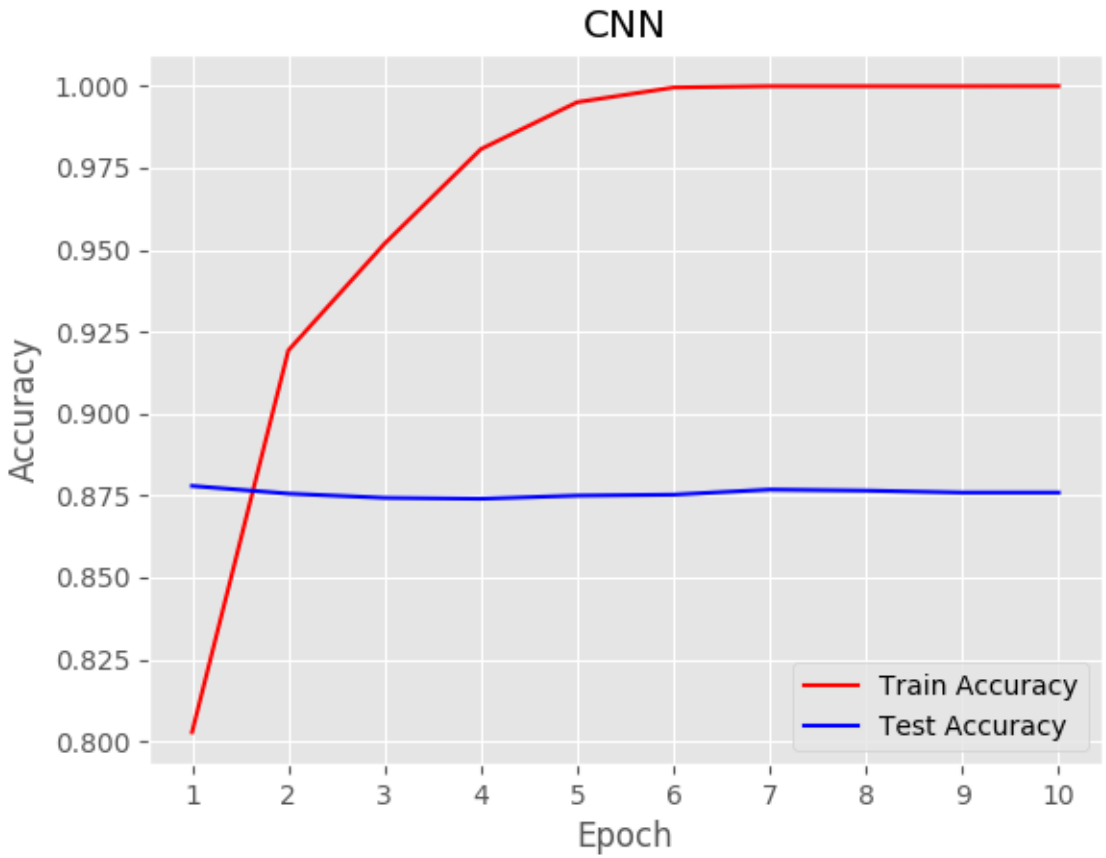
- Good candidate since words have a strong dependencies with previous and next word neighbours.

- Used a lot for text predictions.

- MLP binary classification as last layer.



4. Results - CNN



5. Discussion

- For the three models, accuracy greater than 85 % achieved.
- CNN is the best approach in the task, with an accuracy of 87.59 % after 10 epochs.
- Networks might be too big.
- LSTM costly trained. Not worth it.
- Kaggle Competition → Position 267 / 578

6. Conclusion

- Indeed good classification results.
- CNN does not add a significant better result than a MLP.
- Similar performance of models

Open Questions:

- Check for objective and subtle reviews
- Apply dropout not to overfit the networks

The End

Thank you for your attention.
Any question?